

DNA Binding Specificity of the Mu Ner Protein<sup>†</sup>Teresa E. Strzelecka,<sup>‡</sup> Jeffrey J. Hayes,<sup>§||</sup> G. Marius Clore,<sup>\*,‡</sup> and Angela M. Gronenborn<sup>\*,‡</sup>*Laboratory of Chemical Physics, NIDDK, and Laboratory of Molecular Embryology, NICHD, National Institutes of Health, Bethesda, Maryland 20892**Received July 22, 1994; Revised Manuscript Received December 12, 1994<sup>⊗</sup>*

**ABSTRACT:** Binding of purified phage Mu Ner protein to a series of DNA fragments was investigated in order to determine the length requirements for tight specific binding. Gel retardation experiments with wild-type 307 base pair (bp) Mu DNA and shorter, synthetic oligonucleotides were performed, and apparent dissociation constants ( $K_D^{\text{app}}$ ) were determined from the half-saturation point. While Ner formed four complexes with the 307 bp DNA fragment, only one complex was observed with the shorter DNAs. The 50 and 30 bp fragments had  $K_D^{\text{app}}$  values of 5 and 20 nM, respectively. Ner binding was progressively weaker with decreasing size of the DNA fragments, with no binding observed for 12mers. The shortest DNA fragments which bound well were two 18 bp fragments for which  $K_D^{\text{app}}$  values were in the range of 50–100 nM. The stoichiometry of Ner complexes with the 30 and 18 bp fragments was determined using a modified Ferguson method. Ner was found to form a tetramer on the 30 bp DNA and a dimer on the 18 bp DNA, which makes the latter a good candidate for the study of a Ner–DNA complex by NMR. In order to clarify which DNA regions were important for Ner–DNA binding, hydroxyl radical footprinting was performed for a range of Ner concentrations from 30 to 500 nM. The footprint revealed that Ner contacts the DNA backbone every 12–13 bp, on both strands of the DNA. The order in which protected regions appeared with increasing protein concentration indicated that two Ner monomers bound to DNA simultaneously. A model of Ner binding to DNA is proposed on the basis of these results.

The Ner protein is a small (74 amino acids), basic DNA binding protein which regulates a switch between the lysogenic and lytic life cycles of phage Mu (van de Putte *et al.*, 1980). When bound to DNA, Ner prevents transcription of the *c* repressor protein which maintains the lysogenic state and also down-regulates its own transcription (van Leerdam *et al.*, 1982).

Mu Ner is an interesting candidate for structural studies for several reasons. There are only two other proteins which have significant amino acid sequence homology with Ner: the Ner protein of phage D108, which also regulates the lysogenic/lytic switch (Hull *et al.*, 1978; Gill *et al.*, 1981), and the Ner-like protein, Nlp, of *Escherichia coli* (Choi *et al.*, 1989). In the conserved 63 amino acid region, there is 53% sequence identity between D108 Ner and Mu Ner and 65% identity between Nlp and Mu Ner. Despite the high level of sequence homology, the two Ner proteins do not bind to each other's DNA sites. D108 Ner binds to a GC-rich 47 bp DNA fragment containing two perfect 11 bp inverted repeats separated by an 8 bp spacer (Kukulj *et al.*, 1989b). The *nlp* gene, isolated from *E. coli* K-12 strain W3110, encodes a 91 amino acid protein which activates *mal* gene expression and was found to be nonessential for *E. coli* viability (Autexier & DuBow, 1992). Its DNA binding site is not currently known.

The secondary structure of the Mu Ner protein in solution has been determined by two-dimensional NMR (Gronenborn

*et al.*, 1989). It shows that Ner is an  $\alpha$ -helical protein. Despite functional similarities, however, between the cro protein of  $\lambda$  phage and Mu Ner, there is no obvious amino acid homology between the two proteins. Further, no homology was found with prokaryotic proteins which employ the helix–turn–helix (HTH) motif in DNA recognition ( $\lambda$  repressor, 434 cro, etc.). The  $\alpha$ -helical regions of Ner, on the other hand, overlap with the clusters of amino acids conserved between the two Ner proteins and Nlp, indicating that the overall structures of these proteins may be similar.

To date, no three-dimensional structure of Mu Ner has been determined, on the other hand, partly because of assignment problems due to severe overlap of hydrogen resonances. Using heteronuclear 3D and 4D NMR techniques [for reviews, see Clore and Gronenborn (1991) and Bax and Grzesiek (1993)], this problem can now be overcome, and a structure determination of Ner is underway. In parallel, we initiated a study of Ner binding to DNA in order to evaluate whether the determination of Ner–DNA complex structure by NMR was feasible. The Ner binding site on Mu DNA was identified by genetic (Goosen & van de Putte, 1984) and DNase I footprinting experiments (Tolias & DuBow, 1986; Kukulj *et al.*, 1989a). It comprises a 50 bp AT-rich fragment containing two 12 bp imperfect inverted repeats separated by a 6 bp spacer. This fragment overlaps with both the repressor promoter ( $P_c$ ) and the early promoter ( $P_e$ ), from which the *ner* gene is the first transcribed (Figure 1). Since only DNase I footprinting data were available for Mu Ner, binding constants for different sizes of DNA fragments had to be determined in order to delineate the smallest possible DNA site for stable complex formation, and information about the stoichiometry of complexes had to be obtained. Results of these experiments are presented in this paper.

<sup>†</sup> This work was supported by the AIDS-Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health (to G.M.C. and A.M.G.).

\* Authors to whom correspondence should be addressed.

<sup>‡</sup> NIDDK.

<sup>§</sup> NICHD.

<sup>||</sup> Present address: Department of Biochemistry, University of Rochester, Rochester, NY 14642.

<sup>⊗</sup> Abstract published in *Advance ACS Abstracts*, February 1, 1995.

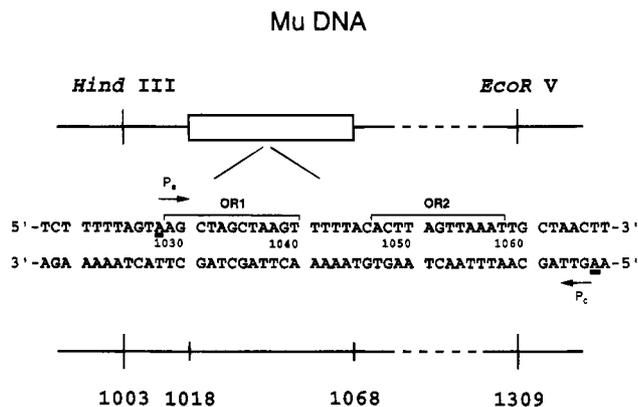


FIGURE 1: Location of the 307 bp fragment and Ner binding site on Mu DNA. Boxes outline the OR1 and OR2 operator regions proposed by Kukolj *et al.* (1989a). Positions of the transcription starts from  $P_e$  and  $P_c$  promoters are underlined, with arrows indicating the direction of transcription.

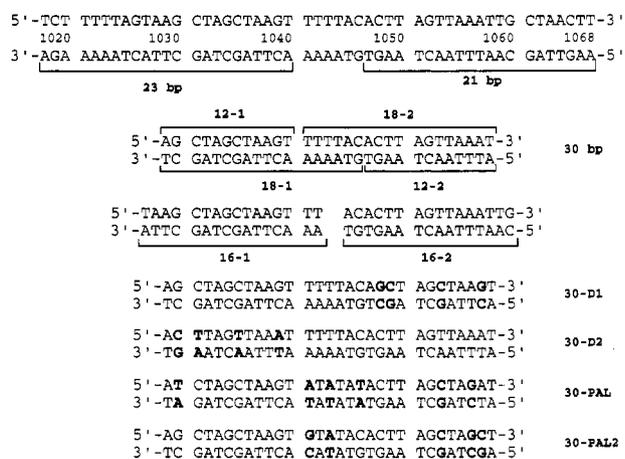


FIGURE 2: Sequences of the 12–50 bp fragments used in the gel retardation experiments. In the symmetrized fragments (30-D1 to 30-PAL2), bases which have been changed with respect to the original 30mer are shown in boldface.

## MATERIALS AND METHODS

**Protein Samples.** Ner was purified from *E. coli* B cells containing the expression plasmid pL-ner as described previously (Allet *et al.*, 1988). Ner concentrations were determined using the absorbance value  $A^{1\%} = 23.2$  at 280 nm (Allet *et al.*, 1988). Approximately 10% of the Ner molecules were found to have the pyruvate-modified cysteine (Rose *et al.*, 1992). Samples used in the gel retardation experiments were dialyzed and concentrated against 50 mM potassium phosphate/0.15 M KCl, pH 7.0, using a Micro-ProDiCon device (Spectrum). Subsequent protein dilutions were made in the same buffer.

**DNA Fragments.** The 12–50 bp DNA fragments and the PCR primers were synthesized with an Applied Biosystems 380B DNA synthesizer. Sequences of these fragments are shown in Figure 2. The 12–23 bp fragments and primers were purified by HPLC (Pharmacia–LKB) using the Dynamax 300A column (Rainin). The 30 and 50 bp fragments used for gel retardation experiments were purified on 18% acrylamide gels. DNA was eluted at 37 °C overnight in elution buffer (0.75 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, and 0.1% SDS) and precipitated with ethanol. Pellets obtained after centrifugation at 15 000 rpm for 90 min were dried and dissolved in H<sub>2</sub>O, and the DNA solutions were desalted using EconoPac 10DG columns (Bio-Rad).

One strand of each duplex was end-labeled with [ $\gamma$ -<sup>32</sup>P]-ATP (Amersham) using T4 DNA kinase (New England Biolabs) according to Maniatis *et al.* (1982). Unincorporated label was removed using the G-25 Aqua-Select columns (5 Prime → 3 Prime, Inc.). Labeled duplexes were obtained by mixing complementary strands in equimolar amounts in 0.1 M KCl, heating to 95 °C, and slowly cooling to room temperature in a water bath.

To check the binding of purified Ner to wild-type DNA, a 307 bp DNA fragment was prepared by PCR from a pUC9-ner plasmid (the *Hind*III–*Hinf*I fragment of Mu DNA, base pairs 1002–1360, was cloned into the pUC9 *Hind*III site; unpublished data). The Ner binding site is located on the left side of this fragment (see Figure 1). The primers used were 5'-CTTTACATTAAGCTTTTCAGTAATTAT-3' (*Hind*III site underlined) and 5'-CTGCTTGATATCGGCT-AGGCC-3' (*Eco*RV site underlined). All restriction enzymes used were from New England Biolabs. The PCR template was a pUC9-ner plasmid linearized by cutting with *Hind*III and purified using the Qiagen tip-20 and protocol for DNA purification after enzymatic modification (Qiagen, Inc.). After PCR, the fragment was cut with *Eco*RV and *Hind*III, dephosphorylated using 0.1 unit of calf intestinal phosphatase (CIP, New England Biolabs) per picomole of DNA ends, purified using the Qiagen protocol, and precipitated with ethanol. Dried pellets were dissolved in H<sub>2</sub>O and desalted on G-50 Select columns (5 Prime → 3 Prime). DNA was end-labeled with [ $\gamma$ -<sup>32</sup>P]ATP using T4 nucleotide kinase and the unincorporated label removed on G-50 Select columns. The 301 bp fragment used for competition experiments contains the *E. coli* lacZ region, bp 1008–1308 (Hardies *et al.*, 1979).

**Gel Retardation Experiments.** Ner was freshly diluted for each assay. The binding buffer contained 40 mM Tris-HCl, pH 8.0, 0.15 M KCl, 5 mM MgCl<sub>2</sub>, 0.01–0.03  $\mu$ g/ $\mu$ L BSA (bovine serum albumin), and 1 mM DTT. Concentrations of labeled DNA fragments ranged from 5 to 500 pM. The protein was incubated with DNA in 1.5 mL microcentrifuge tubes for 30 min at 37 °C. A 0.1 volume of a loading buffer (50% glycerol/0.02% xylene cyanol) was added to the samples after incubation. Polyacrylamide gels [acrylamide:bis(acrylamide) ratio 37.5:1] cast with 1× TBE buffer (9 mM Tris-HCl, 8.8 mM boric acid, and 2 mM EDTA) were prerun in 1× TBE for 1 h at 200 V. Samples were loaded at 50 V. The gels were run at 250 V until samples entered the gel, and then at 200–220 V for 2–3 h.

After the gels were dried, they were exposed to Kodak XAR 5 film at –70 °C using intensifying screens. To quantify the intensities of free and bound DNA bands, gels were placed in phosphorimager cassettes, and screens were exposed from 1 to 12 h. The screens were scanned using a Molecular Dynamics PhosphorImager 400 with the Image Quant software. Three scans were performed for each lane and their results averaged. The fraction of bound (free) DNA at each protein concentration was calculated by dividing the area of bound (free) bands by the total area of bound and free bands.

**Determination of the Stoichiometry of Ner–DNA Complexes.** To determine the stoichiometry of Ner–DNA complexes, both a simple gel retardation assay and a modified version of the Ferguson method (Ferguson, 1964; Bryan, 1977) were used. The Ferguson method was originally designed to determine molecular weights of proteins in native gels. Recently it has also been used to determine

molecular weights of protein–DNA complexes (May *et al.*, 1991; Orchard *et al.*, 1992; Orchard & May, 1993).

In a gel retardation assay, unlabeled 30 bp DNA fragments (0.5  $\mu\text{g}$ ) and varying Ner:DNA ratios were used. Samples of protein–DNA complexes were prepared in the same way as for labeled DNA. Samples were loaded onto a 12% acrylamide gel and run at 220 V for 2.5 h. The gel was stained in 0.5  $\mu\text{g}/\text{mL}$  ethidium bromide for 20 min and photographed. For the Fergusson analysis, a set of six acrylamide gels with concentrations of 7, 8, 9, 10, 11, and 12% were cast. Each gel contained 0.25 $\times$  TBE and 2.5% sucrose. Ten micrograms of each protein standard (non-denatured molecular weight markers, Sigma) and Ner + 18 bp (18-1), Ner + 30 bp, and *Bam*HI + 12 bp complexes were also loaded onto each gel. The *Bam*HI + 12mer complex was used as a test of the accuracy of this method, since its stoichiometry is known (2:1 protein:DNA ratio). *Bam*HI and the 12mer were a gift from Dr. Aneel Aggarwal, Columbia University. The protein–DNA complexes were formed at protein concentrations below 30  $\mu\text{M}$  and subsequently concentrated with buffer exchange. The final buffer concentration in these samples was approximately 5 mM KCl/10 mM potassium phosphate, pH 7.0. The amount of protein in each of the complexes loaded onto a gel was 3.9  $\mu\text{g}$  for Ner + 18 bp, 6  $\mu\text{g}$  for Ner + 30 bp, and 3.5  $\mu\text{g}$  for *Bam*HI + 12 bp complex.

Gels were prerun in 0.25 $\times$  TBE for 1 h at 100 V, and then run at 230 V until samples entered the gel and at 200 V until the bromophenol blue bands were about 1 cm from the gel bottom. The distances of these bands from the wells for each lane were measured, and the gels were stained with Coomassie blue overnight. After the gels were destained, the distance of each protein and each protein–DNA complex band from the wells was measured. The relative mobility of each band ( $R_f$ ) was calculated by dividing its distance by the distance of the bromophenol blue band in the same lane. Since some of these protein standards contain more than one band due to the presence of charge isomers, the  $R_f$  of only the major band was used in data analysis. The plots of 100-[log(100 $R_f$ )] vs gel concentration were constructed for each protein standard and for the protein–DNA complexes. The slopes of these plots (retardation coefficients) were then plotted against the molecular weights of protein standards on a log–log or linear scales, and these calibration lines were used to determine the sizes of protein–DNA complexes.

**Hydroxyl Radical Footprinting.** The DNA fragments used for hydroxyl radical footprinting were obtained by PCR amplification of the 991–1095 bp region of a pMu6 plasmid (Priess *et al.*, 1982) which contains the 50 bp Ner binding site in its center. The primers were 5'-GCTCTAGAT-TAAGCTTTTCAGT-3', *Xba*I site underlined, and 5'-AA-GATCCTCCTAAGTTTGTAAAT-3', *Mbo*I site underlined. This fragment was cloned into the pT7Blue vector (Novagen), and two of the clones were sequenced by standard methods. DNA from two clones of correct sequence was cut with *Spe*I and *Eco*RI or *Xba*I and *Eco*RI, respectively, and labeled using [ $\alpha$ - $^{32}\text{P}$ ]dCTP and Klenow fragment of DNA polymerase. The *Spe*I–*Eco*RI fragment (157 bp) was labeled at the top strand and the *Xba*I–*Eco*RI fragment (169 bp) at the bottom strand of the Ner sequence. Following labeling, fragments were purified on an 8% acrylamide gel, eluted into TE buffer (10 mM Tris-HCl/1 mM EDTA) overnight, precipitated twice with ethanol, and dissolved in TE to a final concentration of  $3 \times 10^3$  cpm (estimated DNA concentration was 10–25 fmol/ $\mu\text{L}$ ).

Ner–DNA complexes were prepared by adding Ner dilutions to 10  $\mu\text{L}$  of labeled DNA in a binding buffer which was the same as that for the gel retardation assays except that the Tris-HCl concentration was lowered to 10 mM to avoid possible interference with the hydroxyl radical reaction. The final volume of the binding reactions was 35  $\mu\text{L}$ . Following incubation at room temperature for 30 min, the hydroxyl radical cleavage reaction was initiated by addition of 5  $\mu\text{L}$  from the 10 $\times$  stock solutions of 1 mM [FeEDTA] $^{2-}$ , 10 mM sodium ascorbate, and 0.125% hydrogen peroxide (Dixon *et al.*, 1991). After 1 min at room temperature, the reaction was stopped by adding 50  $\mu\text{L}$  of a stop solution (0.05 M thiourea, 0.6 M sodium acetate, 5 mM EDTA, and 0.4% SDS). DNA was precipitated with ice-cold ethanol, washed with 70% ethanol, dried, and dissolved in 20  $\mu\text{L}$  of TE; 5  $\mu\text{L}$  aliquots were dried, dissolved in 4  $\mu\text{L}$  of sample buffer (100% formamide, 2 mM EDTA, 0.02% of bromophenol blue, and xylene cyanol), and run on 8% sequencing gels. After the gels were dried, they were exposed to Kodak XAR 5 film at room temperature for 3–4 days. Two-dimensional scans of exposed films were obtained on a Molecular Dynamics Computing Densitometer at 100 nm resolution.

## RESULTS

**Determination of Equilibrium Dissociation Constants.** Binding of purified Ner to DNA was examined first for the “wild-type” 307 bp fragment obtained by PCR from the pUC9-ner plasmid. Gel retardation assays using this DNA were performed in the presence of both specific (same fragment, unlabeled) and nonspecific DNA (301 bp DNA). The results are shown in Figure 3. As can be seen in Figure 3A, the first Ner–DNA complex is formed at a Ner concentration of about 10 nM, and three more complexes are formed when the protein concentration is increased up to 500 nM. No further shift of a DNA band can be observed up to a Ner concentration of 10  $\mu\text{M}$ . To determine whether all four complexes are specific, two competition experiments with the unlabeled 307 bp fragment were performed, at Ner concentrations of 10 nM (Figure 3B) and 500 nM (Figure 3C). Formation of complexes can be suppressed by the same DNA. A competition experiment with unlabeled 301 bp fragment at 500 nM Ner showed that Ner binding in complex 4 is at least 1000 times more specific than binding to this fragment, since no decrease in the intensity of the slowest migrating band was observed up to 100 nM of 301 bp DNA (Figure 3D).

Binding of Ner to the 50 and 30 bp synthetic oligonucleotides is shown in Figure 4A and Figure 4B, respectively. The stoichiometry of the Ner complex with the 30 bp fragment is 4:1 (see below). The isotherms for the binding of Ner to DNA, describing the dependence of the fraction  $\phi$  of DNA that is bound as a function of the free protein concentration  $P$ , will then be given by

$$\phi = KL/(1 + KL) \quad (1)$$

if Ner is a tetramer in solution, by

$$\phi = K_{\text{app}}^2 L^2 / (1 + K_{\text{app}}^2 L^2) \quad (2)$$

if Ner is a dimer in solution, and by

$$\phi = K_{\text{app}}^4 L^4 / (1 + K_{\text{app}}^4 L^4) \quad (3)$$

if Ner is a monomer in solution (Clare *et al.*, 1982). In the

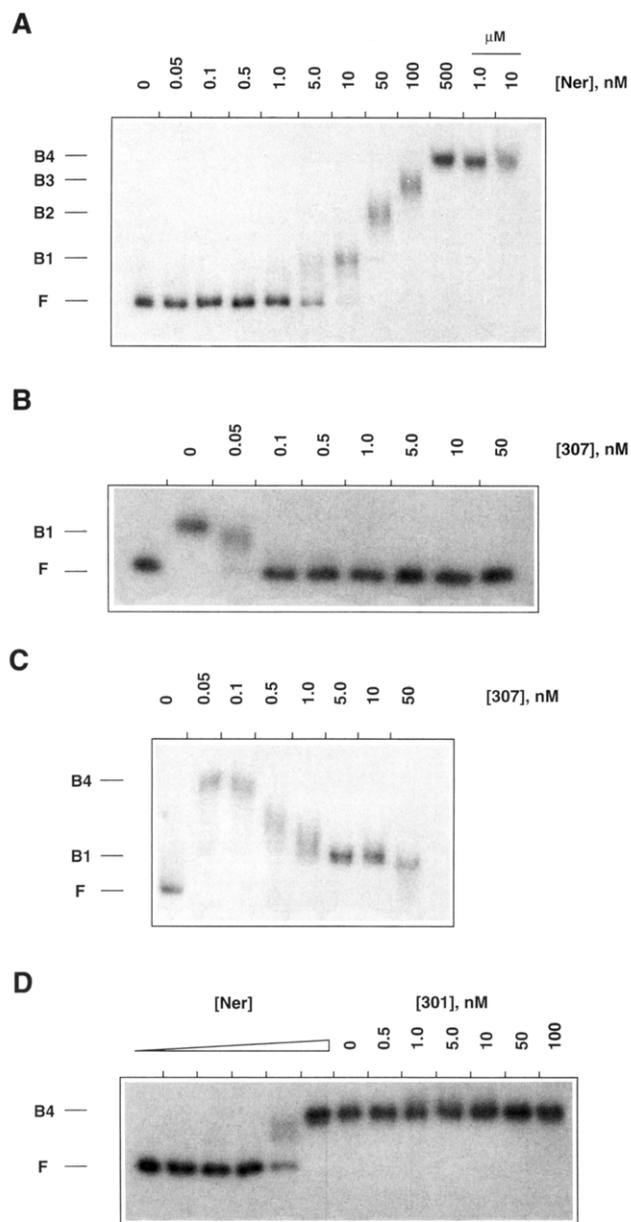


FIGURE 3: Binding of Ner to the 307 bp Mu DNA fragment. (A) No competing DNA,  $[307^*] = 40$  pM (asterisk indicates radiolabeled DNA); (B) 10 nM Ner,  $[307^*] = 50$  pM in the presence of increasing concentrations of unlabeled 307 bp fragment; (C) 500 nM Ner,  $[307^*] = 10$  pM in the presence of increasing concentrations of unlabeled 307 bp fragment; (D)  $[307^*] = 100$  pM in the presence of increasing amounts of unlabeled 301 bp fragment. Ner concentrations are the following: lane 1, 0 nM; 2, 5 nM; 3, 10 nM; 4, 50 nM; 5, 100 nM; 6–13, 500 nM. Gels A–C were 6% acrylamide; gel D was 8% acrylamide. F, free DNA; B1–B4, Ner–DNA complexes one to four.

latter two cases, binding of Ner to DNA must be highly cooperative as no binding intermediates are observed for either the 30 or the 50 bp fragments (Figure 4). Only eq 2, that is, cooperative binding of two dimers, could fit the experimental data adequately (Figure 5), yielding a value of  $K_{app} = 5.2 (\pm 0.5) \times 10^7 M^{-1}$ . This value of  $K_{app}$  subsumes a cooperativity parameter  $\alpha$  where the true equilibrium association constant  $K$  is given by  $K_{app}/\sqrt{\alpha}$ . (Note that  $K$  and  $\alpha$  cannot be determined independently from the experimental data in Figure 5.)

Apparent equilibrium dissociation constants ( $K_D^{app}$ ) were determined from the half-saturation points for at least two gels in the case of 16-1, 16-2, 18-1, 18-2, 21, 23, 30, and 50

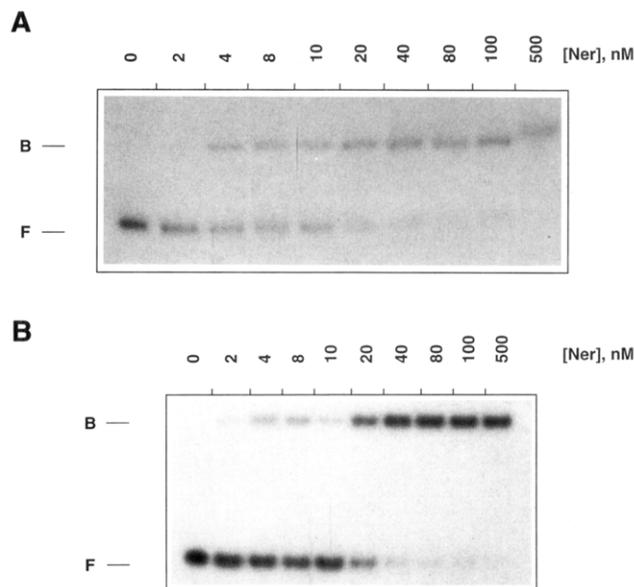


FIGURE 4: Ner binding to the 50 bp (A) and 30 bp (B) synthetic oligonucleotides. Both gels were 12% acrylamide;  $[50^*] = 100$  pM,  $[30^*] = 100$  pM. F, free DNA; B, Ner–DNA complex.

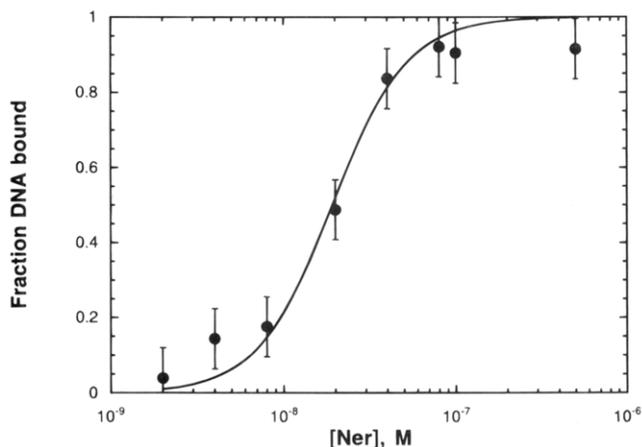


FIGURE 5: Fraction of bound DNA ( $\theta$ ) in the Ner + 30 bp complex vs protein concentration for the gel from Figure 4B. The dashed line represents the best fit to eq 2 with  $K_{app} = 5.2 (\pm 0.5) \times 10^7 M^{-1}$ .

Table 1: Apparent Equilibrium Dissociation Constants for Ner–DNA Binding Determined from Gel Retardation Assays

oligonucleotide	$K_D^{app}$	oligonucleotide	$K_D^{app}$
12-1, 12-2	$> 10 \mu M$	23	25 nM
16-1, 16-2	$\sim 500$ nM	30	20 nM
18-1, 18-2	50–100 nM	50	5 nM
21	30 nM		

bp fragments and are listed in Table 1. No binding was observed for the two 12 bp fragments for Ner concentrations up to  $10 \mu M$ .  $K_D^{app}$  generally increased as the size of the DNA fragments decreased. However, there was at least a 5-fold increase in binding between the 16 and 18 bp fragments, indicating that bases in the 6 bp spacer between OR1 and OR2 may be important for specific Ner–DNA interactions. Indeed, subsequent hydroxyl radical footprinting showed that the DNA backbone at some of these positions is protected by the protein from hydroxyl radical attack (see below).

*Hydroxyl Radical Footprinting.* The hydroxyl radical footprints for the bottom and top strands of the Ner binding

Table 2: Molecular Weights of Ner–DNA and *Bam*HI–DNA Complexes Obtained from the Ferguson Method<sup>e</sup>

complex	(–slope)	MW <sup>a</sup>	MW <sup>b</sup>	MW <sup>c</sup>	MW <sup>d</sup>	<i>n</i> <sup>a</sup>	<i>n</i> <sup>b</sup>	<i>n</i> <sup>c</sup>	<i>n</i> <sup>d</sup>
<i>Bam</i> HI + 12 bp	5.03 (5.7%)	56780 (6.9%)	53870 (12.6%)	59360 (12.3%)	55550 (23.5%)	1.99	1.87	2.09	1.94
Ner + 30 bp band 1	4.79 (2.4%)	54560 (3.6%)	51490 (9.3%)	57080 (9.0%)	52730 (20.2%)	4.09	3.73	4.39	3.87
Ner + 30 bp band 2	6.31 (2.4%)	68280 (3.6%)	66210 (9.3%)	70740 (9.0%)	70220 (20.2%)	5.70	5.46	5.99	5.93
Ner + 18 bp	3.21 (6.9%)		36260 (13.8%)		34860 (24.7%)		2.86		2.70

<sup>a</sup> Determined using the (–slope) vs MW calibration, no  $\alpha$ -lactalbumin. <sup>b</sup> Same as *a*, but with  $\alpha$ -lactalbumin. <sup>c</sup> Determined using the log(–slope) vs log(MW) calibration, no  $\alpha$ -lactalbumin. <sup>d</sup> Same as *c*, but with  $\alpha$ -lactalbumin. <sup>e</sup> Estimated molecular weights of the protein–DNA complexes are as follows: *Bam*HI dimer + 12 bp, 57 060; Ner dimer + 18 bp, 28 880; Ner tetramer + 30 bp, 53 800; Ner tetramer + 2  $\times$  30 bp, 73 600.

site are shown in Figure 6A. Densitometer scans of the gels for the top and bottom strands are shown in Figure 6B,C, respectively. Superposition of these results onto the DNA sequence (Figure 6D) shows that at 30 nM Ner only backbone positions within the 30 bp region containing OR1 and OR2 are protected. With further increases in protein concentration, additional protected sites appear around bases 20, 30, 80, and 90. Closer inspection of densitometer scans reveals an order in which these appear. A schematic diagram of site protection with increasing Ner concentration is shown in Figure 7. At 30 nM Ner, strong protection is observed on the top strand centered at bases 43–44, 56–57, and 69–70 (sites 2, 1, and 3, respectively). Phosphates protected on the bottom strand at this protein concentration are 41–42, 54–55, and 67–68 (sites 2', 1', and 3'). Spacing between centers of these protected regions on the same strand is 13 bp. There are also four more regions which are weakly protected, with centers at bp 33–34 and 82–83 (top strand, sites 4 and 5), and at bp 29–30 and 80 (bottom strand, sites 4' and 5'). Centers of sites 2 and 4 are separated by 10 bp, and of 2' and 4' by 11 bp, whereas sites 3 and 5 and sites 3' and 5' are separated by 13 bp.

When the concentration of Ner reaches 60 nM, sites 4 and 4' are still weakly protected, whereas 5 and 5' become strongly protected. As the protein concentration increases to 120 nM, sites 4 and 4' become fully protected, with the center of site 4 shifting to bp 32–33. In addition, weakly protected sites 6 (top, bp 19) and 7' (bottom, bp 91) appear. Centers of 4 and 6 are spaced by 13 bp, whereas those of 5' and 7' by 11 bp. At 500 nM Ner, sites 6 and 7' are fully protected, and a weakly protected site 9' at bp 100–101 appears. Spacing between 5' and 7' is 12 bp, and that between 7' and 9' is 9 bp. We have not detected any protection for a region around 20 bp on the bottom strand and around 90 on the top strand.

This footprint has a striking symmetry not only with respect to strand reversal but also in the order of appearance of protected bases. It closely resembles the hydroxyl radical footprint of D108 Ner (Kukolj *et al.*, 1989b). For detailed analysis, see Discussion.

**Determination of the Stoichiometry of Ner–DNA Complexes.** Gel retardation assays using unlabeled 30 bp fragment indicated that four molecules of Ner are present in this complex, since all the DNA was shifted at a 4:1 protein:DNA ratio. To confirm this result, we used the Ferguson method, which was developed for the determination of the molecular weights (MW) of proteins in native gels. This method can also be applied to protein–DNA complexes if the DNA fragments are not too long, so that the complex shape does not deviate significantly from globular. We have found that the most reliable results were obtained when the gels contained 2.5% sucrose and the protein–DNA complexes were as concentrated as possible in a low-salt buffer.

The complexes examined were Ner + 18 bp (18-1), Ner + 30 bp, and *Bam*HI + 12 bp. Ner + 30 bp complex formed two bands, with the second migrating slower than the main band. One of the calibrations used to determine the MW of the Ner + 30 and *Bam*HI + 12 bp complexes is shown in Figure 8. The original Ferguson method used a calibration in the form log(–slope) vs log(MW), whereas Hedrick and Smith (1968) used a (–slope) vs MW relationship. We found that more accurate results were obtained using the Hedrick and Smith (HS) calibration. Including  $\alpha$ -lactalbumin (MW = 14 200) in the calibration resulted in a much larger error of the MW of the complexes. In addition, the band formed by the Ner + 18 bp complex was more diffuse than the other bands.

The estimated sizes of protein–DNA complexes obtained from four different calibrations are shown in Table 2. The relative errors were calculated by adding the relative errors in (–slope) for each of the complexes to the relative error of the slope of the calibration line and are listed in parentheses. Here, *n* equals the number of protein monomers obtained after subtracting the MW of the DNA duplex from the MW of the complex and dividing the result by the MW of the protein monomer. The MW of Ner was assumed to be 8500 and of *Bam*HI 24 570 (Xu & Schildkraut, 1991). Molecular weights of DNA duplexes were 7920, 11 880, and 19 800 for the 12mer, 18mer, and 30mer, respectively. As can be seen from comparison of columns MW<sup>a</sup> and MW<sup>c</sup>, using the log–log calibration results in apparently larger molecular weights and introduces larger errors. Including  $\alpha$ -lactalbumin increases the relative error of MW determination to 9–14%, which is still acceptable considering the uncertainties involved in every step of this process.

Our test of the method's accuracy was the determination of the MW for *Bam*HI + 12mer complex in which two protein monomers are bound per DNA molecule, as determined from biochemical (Nardone & Chirikjian, 1987) and X-ray crystallographic (Newman *et al.*, 1994) studies. The extrapolated values of the number of *Bam*HI monomers in the complex are very close to two. We can therefore conclude with confidence that Ner binds to the 30 bp fragment as a tetramer. The error involved in the estimate of the MW of Ner + 18 bp complex is much larger, but it is clear that Ner is likely to be bound as a dimer to this fragment. The second band in the Ner + 30 bp complex lane migrated slower than the main band, and its intensity was about 25% of the main band. The MW of this complex is such that if we subtract the DNA contribution there seems to be six Ner monomers bound to it. One possibility is that the complex represented by this band has a different conformation of either the DNA and/or Ner and this causes it to migrate more slowly. Another possibility is that there may be two DNA molecules bound by one Ner tetramer, which would give a total MW for the complex of 73 600.

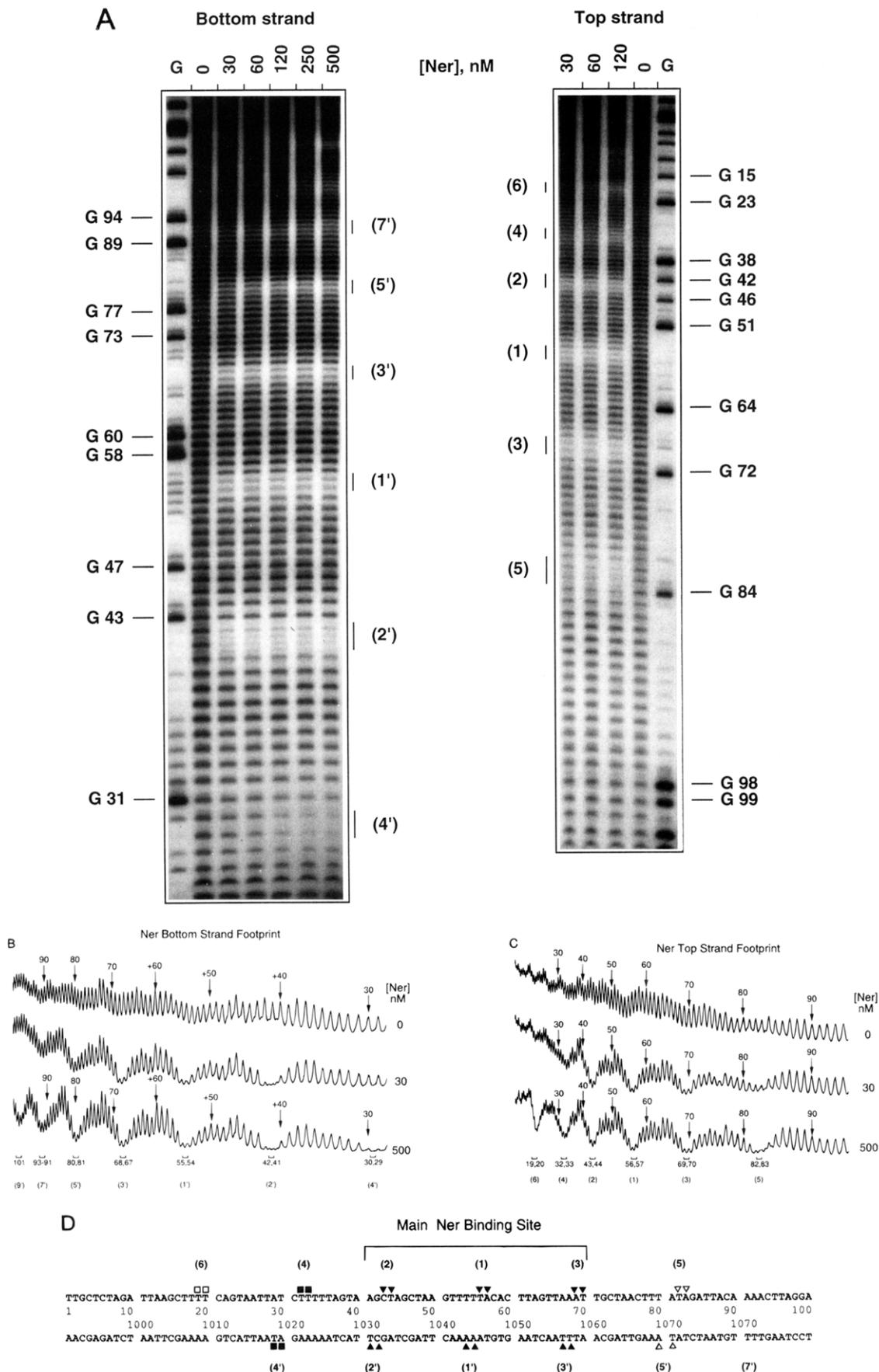


FIGURE 6: (A) Hydroxyl radical footprint of Ner bound to the pMu6-derived DNA fragment labeled on the bottom strand (left) and top strand (right). The leftmost and rightmost lanes show standard Maxam–Gilbert G reactions. The amount of Ner protein added to each reaction is shown above each lane (see Materials and Methods). (B and C) Densitometer scans for the top and bottom strands, respectively. Individual sites of protection are labeled according to their order of appearance with increasing protein concentration (see text). (D) Footprinting results superimposed onto the DNA sequence of the Ner binding site. Centers of sites fully protected by Ner from hydroxyl radical attack are indicated by the following symbols: (▼) 30 nM Ner; (▽) 60 nM; (■) 120 nM; (□) 500 nM.

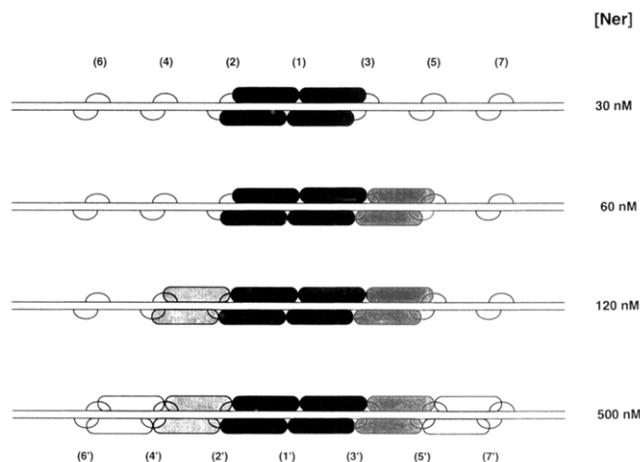


FIGURE 7: Schematic representation of Ner–DNA complex formation with increasing protein concentration deduced from the hydroxyl radical footprinting experiments. The numbers on the top and bottom strands indicate sites protected by Ner from hydroxyl radical attack and are the same as in Figure 6. Shaded ovals represent Ner monomers. The white ovals covering sites 4' and 6' and sites 5 and 7 indicate hypothetical bound monomers, since no protection of sites 6' and 7 was detected by hydroxyl radical footprinting.

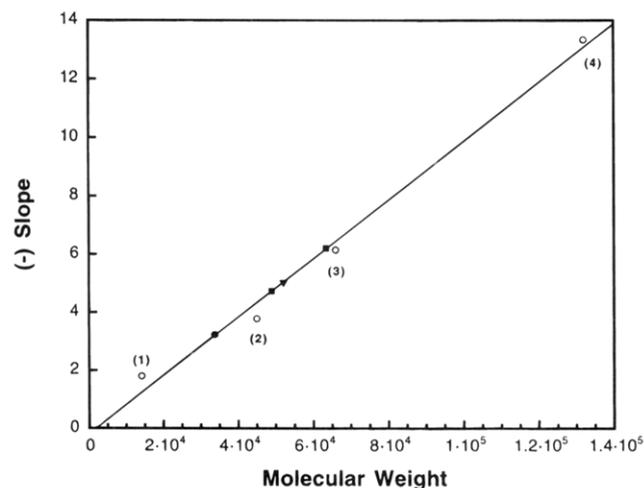


FIGURE 8: Plot of (–slope) vs MW for the following protein molecular weight standards (open circles): (1) carbonic anhydrase, MW = 29 000; (2) chicken egg albumin, MW = 45 000; (3) BSA monomer, MW = 66 000; (4) BSA dimer, MW = 132 000. Protein–DNA complexes are represented by the following symbols: *Bam*HI + 12mer (▼); Ner + 30mer (■); Ner + 18mer (●).

## DISCUSSION

**Ner Binding Site.** Our starting point for the determination of the Ner binding site was the DNase footprint of Mu DNA in the presence of Ner (Kukolj *et al.*, 1989a). It comprises 50 bp of DNA with a stronger protected central 30 bp. The 30 bp fragment, in turn, consists of two 12mers which are imperfect inverted repeats, separated by a 6 bp spacer. This fragment is AT-rich (72% over 50 bp).

Since our goal was to determine the smallest possible binding site with a reasonable binding constant, we had to start by investigating Ner binding to a large DNA fragment containing wild-type Mu DNA. Ner formed four distinct complexes with this DNA over a range of concentrations from 10 to 500 nM, and no additional complexes appeared up to 10  $\mu$ M Ner. The estimated apparent dissociation constant for the first complex is between 5 and 10 nM. All four complexes were specific as can be judged from com-

petition experiments. Binding of Ner to the 50 bp fragment had an estimated  $K_D^{\text{app}}$  of 5 nM, and binding to the 30 bp fragment a  $K_D^{\text{app}}$  of 20 nM, similar to the dissociation constant obtained for the 307 bp DNA. There was only one protein–DNA complex band for these two shorter DNA fragments.

In order to reduce the size even further, we divided the 30 bp fragment into subsites. We started from the two 12mers, for which no binding of Ner was seen up to 10  $\mu$ M protein. Then we tried the 16mers, where two base pairs were included on both sides of the 12mers, and the Ner–DNA complexes were formed with a  $K_D^{\text{app}}$  of about 500 nM. The bands corresponding to Ner–16 bp complexes were quite diffuse, indicating that even within the gel matrix protein was dissociating rapidly from DNA. Ner binding to the 23 and 21 bp fragments was much stronger, but still 3–4-fold lower than binding to the whole 30 bp fragment. It was clear from all these experiments that bases in the six base pair spacer region between the 12mers were important for specific binding of Ner.

This conclusion was supported by experiments where we examined Ner binding to a series of symmetrized 30mers (Figure 2). 30-D1 and 30-D2 had two copies of OR1 and OR2, respectively, resulting in a total of 4 bp changes with respect to the original sequence, whereas 30-PAL and 30-PAL2 had six and five different bases, respectively. 30-PAL and 30-PAL2 which contained altered base pairs within the spacer region bound Ner weaker than 30-D1 and 30-D2 (data not shown).

In order to delineate contacts and boundaries more precisely, it was necessary to carry out footprinting experiments. Single positions along the sugar–phosphate backbone which are contacted closely by a protein can be determined by hydroxyl radical footprinting (Tullius & Dombroski, 1986; Dixon *et al.*, 1991). Even though the footprint was complicated, it was clear that all of the spacer region should be included in the DNA fragment used for binding experiments, since the strongly protected regions 1 and 1' lie within this region. Indeed, binding of the two 18mers, 18-1 and 18-2, was 5–10 times stronger than binding of the 16mers, with 18-1 binding slightly better than 18-2. Moreover, bands corresponding to the Ner + 18 bp complexes were almost as sharp as bands of Ner + 30 bp complexes (on ethidium bromide stained gels), indicating formation of “tight” complexes. Since the molecular weight of the Ner + 18 bp complex is 28 880, determination of its structure by NMR now seems feasible.

**Stoichiometry of Ner–DNA Complexes.** We observed four distinct Ner–DNA complex bands when the 307 bp fragment was used for gel retardation experiments, but only one complex band was present for the 50 bp, 30 bp, and shorter DNA fragments. The number of protein–DNA contacts obtained from the hydroxyl radical footprint indicated that there may be two Ner dimers bound to the 30mer. On ethidium-stained acrylamide gels, all of the 30 bp DNA was shifted at a 4:1 Ner:DNA ratio. We therefore decided to use the Ferguson method to estimate molecular weights of Ner + 30 bp and Ner + 18 bp complexes with the *Bam*HI + 12 bp complex as a standard. When the acrylamide gels were cast without sucrose, the estimated MW of the *Bam*HI + 12 bp complex exceeded 90 000 (should be 57 420), so the method clearly was not working, most likely due to the interaction of proteins and/or DNA with gel matrix. Addition

of 2.5% sucrose produced markedly improved results, even though the Ner + 18 bp complex band was more diffuse than on the gels without sucrose, which resulted in a larger error of its MW estimate. In addition, a second band in the Ner + 30 bp lane was present on gels with sucrose. This extra band was also observed on acrylamide gels of Ner + 30 bp complexes in a very low salt buffer; thus, its appearance is not caused by sucrose in the gel matrix. The molecular weight of this complex was such that it could contain either six Ner molecules bound to one DNA fragment or two DNA molecules bound to a Ner tetramer. The present data are insufficient to distinguish between the two possibilities. However, we do know that a tetramer of Ner is bound to the 30mer and a dimer is bound to the 18mer. Since Mu Ner was found to be monomeric in solution at a concentration of 25  $\mu\text{M}$  (Allet *et al.*, 1988), the dimers are formed in the presence of DNA. Similarly, D108 Ner which is monomeric in solution up to 200  $\mu\text{M}$  dimerized only on DNA (Kukulj *et al.*, 1989b).

**Hydroxyl Radical Footprinting.** At first glance, the hydroxyl radical footprint of Ner within the 30 bp region consisting of OR1, OR2, and a 6 bp spacer closely resembles hydroxyl radical footprints of such helix-turn-helix prokaryotic proteins as  $\lambda$  repressor bound to OR1 (Tullius & Dombroski, 1986) or trp repressor bound to its operator (Carey, 1989), that is until one realizes that there is a Ner tetramer bound to this site. In addition, spacing between centers of protected regions is 13 bp for almost all sites, whereas it is usually 9–10 bp for  $\lambda$  or trp repressors (in the context of hydroxyl radical footprinting, a "protected region" means deoxyriboses within the DNA backbone protected by the protein from hydroxyl radical attack). When mapped onto a B-DNA model with 10.5 bp per turn, protected regions on the same strand would lie roughly on the same side of the DNA helix.

What can we deduce about Ner binding to DNA from this footprint? We know the stoichiometry of the Ner complex with the 30 bp fragment, and thus it should be possible to suggest a plausible arrangement of the protein on the DNA site. A similar pattern of hydroxy radical protection was observed for the Ner protein from the closely related phage D108 (Kukulj *et al.*, 1989b). It consists of five protected regions spanning 56 bp of DNA on each strand, with centers of neighboring regions separated by 11–12 bp on the same strand. It is therefore reasonable to assume that both Ner proteins bind to DNA in a similar manner, despite the fact that their binding sequences are very different. In addition to hydroxy radical footprinting, methylation protection and interference experiments were performed for D108 Ner. The revealed several guanine residues which were protected by D108 Ner from methylation, located between the sites protected from hydroxyl radical attack. Thus, it seems that the sites detected by hydroxyl radical footprinting may correspond to an interface between Ner monomers and/or dimers.

If we assume that this is true for Mu Ner, we can advance the following hypothesis: each pair of hydroxyl radical protected sites on the same strand (i.e., 2 and 1, 1 and 3, etc.) corresponds to a bound monomer, and monomers bound to opposite strands form a dimer. Some of the sites will correspond to interfaces between monomers (or dimers), for example, 1 and 1' for Ner bound to the 30 bp fragment. We also assume that only when a correct Ner–DNA complex is formed do the sites on DNA become fully protected from

hydroxyl radical attack. Confirmation of this hypothesis comes from Ner binding experiments to different length DNA fragments. Let us start with the four complexes observed on the 307 bp fragment, at 10, 50, 100, and 500 nM. At 30 nM Ner, a concentration intermediate between the first and second complexes, we have three strongly protected regions on both strands (2, 1, 3 and 2', 1', and 3') which correspond to a Ner tetramer (see Figure 7). (We also assume that only at a fully protected region is there a correct Ner–DNA complex.) Therefore, the first complex would consist of a Ner tetramer bound to the 307 bp fragment. At 60 nM Ner, just above the concentration when the second complex appears, sites 5 and 5' are strongly protected, indicating that there are two more Ner monomers bound, giving a total of six Ner molecules in the second complex. When the Ner concentration reaches 120 nM, sites 4 and 4' become strongly protected, adding a Ner dimer, for a total of eight Ner molecules in the third complex. Unfortunately, we do not know whether there exists a Ner–DNA complex at a Ner concentration intermediate between 100 and 500 nM. At 500 nM Ner, we also have strongly protected regions 6 and 7'. Since we assumed that addition of Ner molecules proceeds one dimer at a time, it is likely that at this concentration we have a total of 12 Ner molecules bound to DNA.

This hypothesis also works well when we try to rationalize binding of Ner to the shorter DNAs. The 30 bp fragment spans six fully protected sites corresponding to the Ner tetramer. The 23 bp fragment contains sites 2, 2', and 4, and half of 4', whereas the 21 bp DNA contains only sites 3 and 3'. According to our hypothesis, this would mean that two Ner dimers could bind to the 23 bp fragment with two sets of "proper" monomer–monomer contacts, but only one set of such contacts would be present on the 21 bp DNA fragment. Therefore, Ner binding to the 20mers should be weaker than binding to 30mer, and binding to the 21 bp fragment weaker than binding to the 23 bp DNA. Inspection of the apparent dissociation constants from Table 2 shows that this is indeed the case. 18-1 and 18-2, which contain four protected sites, each bind 5–10-fold weaker than the 30mer. 16-1 contains only sites 2, 2', and 1', making it impossible for the protein to make a full set of phosphate contacts with DNA. Since Ner does bind to the 16mers, however weakly, it indicates that some of the bases between these sites must be important for binding as well (see the next section).

Another possibility for the arrangement of Ner dimers would be such that sites on opposite DNA strands, separated by 11 bp (such as 2 and 1', 1 and 3', etc.), would belong to one dimer, since they would lie on the same face of a DNA helix. However, in this case, sites 2' and 4 or 3 and 5' should be strongly protected at the same Ner concentration, and they are not. Also, if this was the case, binding of the 18mers and 16mers should be equally strong, and it is not. Final resolution of this problem will have to await the structure determination of the Ner–DNA complex.

**Bases Important for Ner Specificity.** Sequence comparison of the DNA fragments used for binding experiments shows that certain bases are conserved in all of these DNA stretches on both strands. Figure 9 shows the 17 bp DNA fragments from both strands aligned so that their 5'-end-protected regions coincide. Numbering of these regions is the same as in Figure 6D. There are seven fragments in which the centers of protected regions are separated by 13 bp, two



FIGURE 9: Sequence alignment of the proposed Ner binding fragments from both DNA strands. Regions protected by Ner from hydroxyl radical cleavage are numbered in the same way as in Figure 6D. Conserved bases T7 and A9 are underlined. Bases in which mutations were found to prevent Ner binding are marked by asterisks on the 2→1 fragment.

where the spacing is 12 bp, and one with a spacing of 11 bp. The regions where Ner–DNA backbone contacts occur do not have any obvious sequence homology except that they consist mostly of A's and T's.

It is immediately clear that the central base (8) is almost always A on the top strand fragments and T on the bottom strand, except for the 4→2 fragment, where it is a G. Base 7 is always a T and base 9 an A, but in the 4→2 DNA base 7 is an A and base 9 a T. Base 10 is a G in five fragments, and base 5 is almost always an A or G. A clue as to which of these bases might be important for specific binding comes from studies of mutations which alter Ner binding to DNA (Goosen & van de Putte, 1984). Three mutations were found which prevented Ner binding to DNA: A to G at bp 1037, A to G at bp 1038, and G to T at bp 1040, which would correspond to bases 8–10 in the 2→1 DNA fragment. It is therefore likely that bases necessary for specific recognition by Ner would be AAG on both strands. Another possibility is that the recognition sequence is CTAAG on both strands.

This would explain why we do not observe binding of Ner to the 4→2 and 4'→2' fragments up to 120 nM Ner, whereas the symmetrically positioned 3→5 and 3'→5' fragments are fully protected by Ner at 60 nM. With the alignment for the 4→2 fragment shown here, A8 on the top strand is replaced by G, A9 by T, and G10 by A, whereas on the bottom strand G10 is replaced by A, which would change four out of six bases important for specific recognition, including a mutation which prevents Ner binding, A8 to G. Even with an alternative alignment of the 4→2 region, with A at position 8, we end up with two even more unfavorable changes, namely, A9 to G and G10 to T. It seems that only in the presence of favorable monomer–monomer and dimer–dimer interactions could binding to this site be established.

**Concluding Remarks.** We have shown that Ner protein of phage Mu binds as a tetramer to its main DNA binding site of 30 bp. It has been previously proposed that this binding site consists of two 12 bp operators separated by a 6 bp spacer (Kukolj *et al.*, 1989a). Our data, however, indicate that the Ner operator might in fact be a 17–19 bp DNA fragment with stretches of A's and T's in the regions where Ner makes contacts with the sugar–phosphate backbone. Bases important for specific recognition are most likely to be AAG, as judged from sequence conservation between all of the proposed Ner operators and mutational studies.

At this point, we do not have enough data to present a detailed model of Ner binding to DNA. However, it is possible that its interaction with DNA is similar to the tandem binding of the trp repressor to its half-operator DNA, as documented in a recent X-ray structure by Lawson and Carey (1993), where there are four repressor dimers bound to two 17 bp DNA molecules, forming a left-handed helix of protein wrapped around DNA. This structure reveals extensive monomer–monomer contacts within and between the dimers.

#### ACKNOWLEDGMENT

We thank Dr. Gary Felsenfeld for the use of his laboratory space and Dr. Kiyoshi Mizuuchi for the use of his equipment, and Dr. Altila Szabo for useful discussions.

#### REFERENCES

- Allet, B., Payton, M., Mattaliano, R. J., Gronenborn, A. M., Clore, G. M., & Wingfield, P. T. (1988) *Gene* 65, 259–268.
- Autexier, C., & DuBow, M. S. (1992) *Gene* 114, 13–18.
- Bax, A., & Grzesiek, S. (1993) *Acc. Chem. Res.* 26, 131–138.
- Bryan, J. K. (1977) *Anal. Biochem.* 78, 513–519.
- Carey, J. (1989) *J. Biol. Chem.* 264, 1941–1945.
- Choi, Y.-L., Nishida, T., Kawamukai, M., Utsumi, R., Sakai, H., & Komano, T. (1989) *J. Bacteriol.* 171, 5222–5225.
- Clore, G. M., & Gronenborn, A. M. (1991) *Science* 252, 1390–1399.
- Clore, G. M., Gronenborn, A. M., & Davies, R. W. (1982) *J. Mol. Biol.* 155, 447–466.
- Dixon, W. J., Hayes, J. J., Levin, J. R., Weidner, M. F., Dombroski, B. A., & Tullius, T. D. (1991) *Methods Enzymol.* 208, 380–413.
- Ferguson, K. A. (1964) *Metabolism* 13, 985–1002.
- Gill, G. S., Hull, R. C., & Curtis, R., III (1981) *J. Virol.* 37, 420–430.
- Goosen, N., & van de Putte, P. (1984) *Gene* 30, 41–46.
- Gronenborn, A., Wingfield, P. T., & Clore, G. M. (1989) *Biochemistry* 28, 5081–5089.
- Hardies, S. C., Patient, R. K., Klein, R. D., Ho, F., Reznikoff, W. S., & Wells, R. D. (1979) *J. Biol. Chem.* 254, 5527–5534.
- Hedrick, J. L., & Smith, A. J. (1968) *Arch. Biochem. Biophys.* 126, 155–164.
- Hull, R. C., Gill, G. S., & Curtis, R., III (1978) *J. Virol.* 27, 513–518.
- Kukolj, G., & DuBow, M. S. (1992) *J. Biol. Chem.* 267, 17827–17835.
- Kukolj, G., Tolias, P. P., & DuBow, M. S. (1989s) *FEBS Lett.* 244, 369–375.
- Kukolj, G., Tolias, P. P., Autexier, C., & DuBow, M. S. (1989b) *EMBO J.* 8, 3141–3148.
- Lawson, C. L., & Carey, J. (1993) *Nature* 336, 178–182.
- Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- May, G., Sutton, C., & Gould, H. (1991) *J. Biol. Chem.* 266, 3052–3059.
- Nardone, G., & Chirikjian, J. G. (1987) in *Gene Amplification and Analysis* (Chirikjian, J. G., Ed.) pp 147–184, Elsevier Science Publisher, New York.

- Newman, M., Strzelecka, T. E., Dorner, L. F., Schildkraut, I., & Aggarwal, A. K. (1994) *Nature* 368, 660–664.
- Orchard, K., & May, G. E. (1993) *Nucleic Acids Res.* 21, 3335–3336.
- Orchard, K., Lang, G., Collins, M., & Latchman, D. (1992) *Nucleic Acids Res.* 20, 5429–5434.
- Priess, H., Kamp, D., Kahman, R., Brauer, B., & Delius, H. (1982) *Mol. Gen. Genet.* 186, 315–321.
- Rose, K., Simona, M. G., Savoy, L.-A., Regamey, P.-O., Green, B. N., Clore, G. M., Gronenborn, A. M., & Wingfield, P. T. (1992) *J. Biol. Chem.* 267, 19101–19106.
- Tolias, P. P., & DuBow, M. S. (1986) *Virology* 148, 298–311.
- Tullius, T. D., & Dombroski, B. A. (1986) *Proc. Natl. Acad. Sci. U.S.A.*
- Van de Putte, P., Giphart-Gassler, M., Goosen, N., Goosen, T., & van Leerdam, E. (1988) *Cold Spring Harbor Symp. Quant. Biol.* 45, 347–353.
- Van Leerdam, E., Karreman, C., & van de Putte, P. (1982) *Virology* 123, 19–28.
- Xu, S.-Y., & Schildkraut, I. (1991) *J. Biol. Chem.* 266, 4425–4429.

BI941684O